



Puzzle-Solving Science: The quixotic quest for units in speech perception

Stephen D. Goldinger¹ and Tamiko Azuma²

1: Department of Psychology, Arizona State University, USA

*2: Department of Speech & Hearing Science, Arizona State University, USA
goldinger@asu.edu, azuma@asu.edu*

Although speech signals are continuous and variable, listeners experience segmentation and linguistic structure in perception. For over 30 years, researchers have asked: What constitutes the basic building-block of speech perception? Is it the raw spectrum, the phoneme, the syllable, or something larger? Research on this topic has been cyclic, with a burst of disappointing activity in the early 1970s, followed by a brief period of disinterest, then a period of renewed investigation. In that time, methods to assess speech units have evolved from basic phoneme or syllable monitoring to a variety of indirect methods, such as cross-modal priming and creation of illusory conjunctions. In addition, strictures on stimulus materials have evolved, conditions that create unstable performance have been identified, and many computational models have been advanced. As a result of such research, we now have a slate of candidates for the basic unit of speech perception, each with its own empirical support. Moreover, the literature can be seen oscillating back and forth between conclusions, depending largely upon the authors' original hypotheses.

In this presentation, we will argue for a different perspective on the question of speech units. Specifically, we will suggest that Grossberg's *adaptive resonance theory* (ART) provides a framework to redirect the question away from competitions between hypothesized units, proposing instead that all potential units are emergent properties of dynamic perceptual processes. Rather than viewing units as fixed nodes in a model architecture, ART only views them as potential, transient states. In particular, units only "exist" in ART when they reach a state of adaptive resonance, a level of perceptual coherence that allows conscious encoding. We will outline several basic principles of ART, with special emphasis on ARTPHONE, a specific model of speech perception (Grossberg, Boardman, & Cohen, 1997). We will focus on three principles bearing directly on the question of speech units, including their tenuous existence in processing, their mutually exclusive nature, and the vicissitudes of attention. Following this review, we will demonstrate an ironic outcome: By essentially invalidating the speech-unit question, ART actually validates many previously-dismissed data on phoneme monitoring and other tasks. Moreover, it provides a coherent framework to accommodate most published findings.

With the framework in place, we will report results from five recent experiments, designed not to isolate specific speech units, but to assess the flexible nature of speech dynamics. In the first three experiments, we assessed the power of social influence to affect the outcome of phoneme versus syllable competitions. The data suggest that the "primacy" of speech units is strongly and symmetrically affected by bottom-up and top-down influences. In the last two experiments, we assessed such symmetric influences by manipulating episodic memory traces. By crossing bottom-up (voice) cues and top-down (prior response) cues, these experiments further suggest that unitization in speech perception reflects an adaptive balance of knowledge sources. Specifically, the apparent primacy of either unit—phoneme or syllable—is strongly affected by the last response required by any specific stimulus.

Taken together, the experiments suggest that sublexical units are products of perceptual dynamics, following the same principles that govern word or sentence perception. The principles of adaptive resonance are quite similar

to those of interactive-activation, and ART may have the same unifying potential originally seen in PDP models. Focusing on speech, it provides an elegant framework to combine bottom-up phonetic perception with top-down episodic memory effects, and may thus accommodate a wide range of seemingly disparate data. In that spirit, we will close by noting other phenomena that may indicate underlying resonance, including resonance that arises between speech and reading. [Supported by NIDCD grant R01-DC04535-02.]

Further reading:

- Goldinger, S.D., Azuma, T., Abramson, M., & Jain, P. (1997) Open wide and say “blah!” Attentional dynamics of delayed naming. *Journal of Memory and Language* 37, 190-216.
- Goldinger, S.D. (1998) Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251-279.
- Grossberg, S., Boardman, I., & Cohen, M. (1997) Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance* 23, 483-503.
- Van Orden, G., & Goldinger, S.D. (1994) Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance* 20, 1269-1291.